Portmanteau Test and Simultaneous Inference for Serial Covariances

Han Xiao

Department of Statistics and Biostatistics
Rutgers University
110 Frelinghuysen Road
Piscataway, NJ 08854, USA
e-mail: hxiao@stat.rutgers.edu

and

Wei Biao Wu

Department of Statistics
The University of Chicago
5734 S. University Ave.
Chicago, IL 60637, USA
e-mail: wbwu@galton.uchicago.edu

Abstract: The paper presents a systematic theory for asymptotic inferences based on autocovariances of stationary processes. We consider nonparametric tests for serial correlations using the maximum (or $\mathcal{L}^\infty$) and the quadratic (or $\mathcal{L}^2$) deviations of sample autocovariances. For these cases, with proper centering and rescaling, the asymptotic distributions of the deviations are Gumbel and Gaussian, respectively. To establish such an asymptotic theory, as byproducts, we develop a normal comparison principle and propose a sufficient condition for summability of joint cumulants of stationary processes. We adapt a blocks of blocks bootstrapping procedure proposed by Künsch (1989) and Liu and Singh (1992) to the $\mathcal{L}^\infty$ based tests to improve the finite-sample performance.

AMS 2000 subject classifications: Primary 60F05, 62M10; secondary 62E20.
Keywords and phrases: Autocovariance, blocks of blocks bootstrapping, Box-Pierce test, extreme value distribution, moderate deviation, normal comparison, physical dependence measure, short range dependence, stationary process, summability of cumulants.

1. Introduction

For a real-valued stationary process $\{X_i\}_{i\in\mathbb{Z}}$, from a second-order inference point of view it is characterized by its mean $\mu = \mathbb{E}X_i$ and the autocovariance function $\gamma_k = \mathbb{E}[(X_0-\mu)(X_k-\mu)]$, $k \in \mathbb{Z}$. Assume $\mu = 0$. Given observations $X_1, \ldots, X_n$, the natural estimates of γ_k and the autocorrelation $r_k = \gamma_k/\gamma_0$ are

$$\hat{\gamma}_k = (1/n)\sum_{i=|k|+1}^{n} X_{i-|k|}X_i \quad \text{and} \quad \hat{r}_k = \hat{\gamma}_k/\hat{\gamma}_0,\ 1-n \le k \le n-1, \tag{1}$$

respectively. The estimator $\hat{\gamma}_k$ plays a crucial role in almost every aspect of time series analysis. It is well-known that for linear processes with *independent and identically distributed* (iid) innovations, under suitable conditions, $\sqrt{n}(\hat{\gamma}_k - \gamma_k) \Rightarrow \mathcal{N}(0, \tau_k^2)$, where $\Rightarrow$ stands for convergence in distribution, $\mathcal{N}(0, \tau_k^2)$ denotes the normal distribution with mean zero and variance τ_k^2. Here τ_k^2 can be calculated by Bartlett's formula (see Section 7.2 of Brockwell and Davis (1991)). Other contributions on linear processes include Hannan and Heyde (1972), Hannan (1976), Hosoya and Taniguchi (1982), Anderson (1991), and Phillips and Solo (1992) etc. Romano and Thombs (1996) and Wu (2009) considered the asymptotic normality of $\hat{\gamma}_k$ for nonlinear processes. As a primary goal of the paper, we study asymptotic properties of the quadratic (or $\mathcal{L}^2$) and the maximum (or $\mathcal{L}^\infty$) deviations of $\hat{\gamma}_k$.

1.1. The $\mathcal{L}^2$ Theory

Testing for serial correlation has been extensively studied in both statistics and econometrics, and it is a standard diagnostic procedure after a model is fitted to a time series. Classical procedures include Durbin and Watson (1950, 1951), Box and Pierce (1970), Robinson (1991), and their variants. For a general account of model diagnostics for time series, see Li (2003). The Box-Pierce portmanteau test uses $Q_K = n\sum_{k=1}^{K} \hat{r}_k^2$ as the test statistic, and rejects if it lies in the upper tail of χ_K^2 distribution. An arguable deficiency of this test and many of its modified versions (for a review see for example Escanciano and Lobato (2009)) is that the number of lags K included in the test is held as a constant in the asymptotic theory. As commented by Robinson (1991):

> "*...unless the statistics take account of sample autocorrelations at long lags there is always the possibility that relevant information is being neglected...*"

The problem is particularly relevant if practitioners have no prior information about the alternatives. The incorporation of more lags emerged naturally in the spectral domain analysis; see among others Durlauf (1991), Hong (1996), and Deo (2000). The normalized spectral density $f(\omega) = (2\pi)^{-1}\sum_{k\in\mathbb{Z}} r_k \cos(k\omega)$ is $(2\pi)^{-1}$ when the serial correlation is not present. Let $\hat{f}(\omega) = \sum_{k=1-n}^{n-1} h(k/s_n)\hat{r}_k \cos(k\omega)$ be the lag-window estimate of the normalized spectral density, where $h(\cdot)$ is a kernel function and s_n is the bandwidth satisfying $s_n \to \infty$, including correlations at large lags, and $s_n/n \to 0$. A test for the serial correlation can be obtained by comparing $\hat{f}$ and the constant function $f(\omega) \equiv (2\pi)^{-1}$ using a suitable metric. In particular, using the quadratic metric and rectangle kernel, the resulting test statistic is the Box-Pierce statistic with unbounded lags. Hong (1996) established that

$$\frac{1}{\sqrt{2s_n}}\left(n\sum_{k=1}^{s_n}(\hat{r}_k - r_k)^2 - s_n\right) \Rightarrow \mathcal{N}(0,1), \tag{2}$$

under the condition that the X_i are iid, which implies that all the r_k here are zero. Lee and Hong (2001) and Duchesne, Li, and Vandermeerschen (2010) studied similar tests in the spectral domain, but using a wavelet basis instead of trigonometric polynomials in estimating the spectral density and henceforth working on wavelet coefficients. Fan (1996) considered a similar problem in a different context and proposed the *adapative Neyman* test and thresholding tests, using $\max_{1\le k\le s_n}(Q_k - k)/\sqrt{2k}$ and $n\sum_{k=1}^{s_n}\hat{r}_k^2 I(|\hat{r}_k|>\delta)$ as test statistics, respectively, where δ is a threshold value. Escanciano and Lobato (2009) proposed to use Q_{s_n} with s_n being selected by AIC or BIC.

Whether the iid assumption in Hong (1996) can be relaxed has been an important and difficult problem. Similar problems have been studied by Durlauf (1991), Deo (2000) and Hong and Lee (2003) for the case that X_i are martingale differences. Recently Shao (2011) showed that (2) is true when $\{X_i\}$ is a general white noise sequence, under the geometric moment contraction (GMC) condition. Since the GMC condition, which implies that the autocovariances decay geometrically, is quite strong, the question arises as to whether it can be replaced by a weaker one. Paparoditis (2000) considered a closely related problem in the spectral domain, and derived the limiting distribution of the integrated squared deviation of the ratio between the periodogram and the true spectral density from one; a distinguished feature is that the underlying process is a dependent linear process. To the best of our knowledge, there has been no results if the serial correlation is present in (2). This paper addresses these questions and substantially generalizes earlier results. We prove that (2) remains true even if some or all of the r_k are not zero. The variance of the limiting distribution now depends on the values of r_k. Our result holds for general stationary processes and allows the autocovariances to decay algebraically, extending the applicability of (2). It also helps to understand the joint behavior of sample autocovariances, and can be used to construct confidence regions. We also consider a closely related problem on the limiting distribution of $\sum_{k=1}^{s_n}\hat{r}_k^2$ when serial correlation is present, which enables us to calculate the asymptotic power of the Box-Pierce test with unbounded lags.

1.2. The $\mathcal{L}^\infty$ Theory

A natural choice is to use the maximum autocorrelation as the test statistic. Wu (2009) obtained a stochastic upper bound for

$$\sqrt{n}\max_{1\le k\le s_n}|\hat{\gamma}_k - \gamma_k|, \tag{3}$$

and argued that in certain situations the test based on (3) has a higher power than the Box-Pierce tests with unbounded lags in detecting weak serial correlation. It turns out that the uniform convergence of autocovariances is also closely related to the estimation of orders of ARMA processes or linear systems in general. The pioneer works in this direction were by E. J. Hannan and his collaborators, see for example Hannan (1974) and An, Chen, and Hannan (1982). For a summary of these works see Hannan and Deistler (1988) and the references therein. In particular, An, Chen, and Hannan (1982) showed that if $s_n = O[(\log n)^\alpha]$ for some $\alpha < \infty$, then with probability one

$$\sqrt{n}\max_{1\le k\le s_n}|\hat{\gamma}_k - \gamma_k| = O\left(\log\log n\right). \tag{4}$$

The question of deriving the asymptotic distribution of (3) is more challenging. Although Wu (2009) was not able to obtain the limiting distribution, his work provided insights into this problem. Assuming $k_n \to \infty$,

$k_n/n \to 0$ and $h \geq 0$, he showed that, for $T_k = \sqrt{n}(\hat{\gamma}_k - \mathbb{E}\hat{\gamma}_k)$,

$$(T_{k_n}\,,\, T_{k_n+h})^\top \Rightarrow \mathcal{N}\left[0, \begin{pmatrix} \sigma_0 & \sigma_h \\ \sigma_h & \sigma_0 \end{pmatrix}\right], \quad \text{where } \sigma_h = \sum_{k\in\mathbb{Z}} \gamma_k\gamma_{k+h}, \tag{5}$$

and we use the superscript $\top$ to denote the transpose of a vector or a matrix. The asymptotic distribution in (5) does not depend on the speed that $k_n \to \infty$. It suggests that, at large lags, the covariance structure of (T_k) is asymptotically equivalent to that of the Gaussian sequence

$$(G_k) := \left(\sum_{i\in\mathbb{Z}} \gamma_i\eta_{i-k}\right) \tag{6}$$

where η_i's are iid standard normal random variables. Let

$$a_n = (2\log n)^{-1/2} \;\text{ and }\; b_n = (2\log n)^{1/2} - (8\log n)^{-1/2}(\log\log n + \log 4\pi). \tag{7}$$

According to Berman (1964) (also see Theorem 14), under the condition $\lim_{n\to\infty} \mathbb{E}(G_0G_n)\log n = 0$,

$$\lim_{s\to\infty} P\left(\max_{1\leq i\leq s} |G_i| \leq \sqrt{\sigma_0}(a_{2s}\,x + b_{2s})\right) = \exp\{-\exp(-x)\}.$$

Wu (2009) conjectured that under suitable conditions, one has the Gumbel convergence

$$\lim_{n\to\infty} P\left(\max_{1\leq k\leq s_n} |T_k| \leq \sqrt{\sigma_0}(a_{2s_n}\,x + b_{2s_n})\right) = \exp\{-\exp(-x)\}. \tag{8}$$

The law with the distribution function $\exp\{-\exp(-x)\}$ is called the *extreme value distribution of type I* or *Gumbel distribution.* In a recent work, Jirak (2011) proved this conjecture for linear processes and for s_n growing with at most logarithmic speed. We prove (8) in Section 4 for general stationary processes, and our result allows s_n to grow as $s_n = O(n^\eta)$ for some $0 < \eta < 1$, with η arbitrarily close to 1, under appropriate moment and dependence conditions. This result substantially relaxes the severe restriction on the growth speed in (4) and in Jirak (2011). The distributional convergence is more useful for statistical inference. For example, other than testing for serial correlation, (8) can be used to construct simultaneous confidence intervals of autocovariances. We also extend (4) for general stationary processes with the maximum taken over the range $1 \leq k < n$. Other than estimating the order of a linear system, the uniform convergence rate is also useful for bandwidth selection of spectral density estimation, see Politis (2003) and Paparoditis and Politis (2012).

1.3. Relations with the Random Matrix Theory

In a companion paper, using the asymptotic theory of sample autocovariances developed here, Xiao and Wu (2012) studied convergence properties of estimated covariance matrices that are obtained by banding or thresholding. Their bounds are analogs under the time series context to those of Bickel and Levina (2008a,b).

There is an important difference between the two settings: we assume only one realization is available, while Bickel and Levina (2008a,b) require multiple iid copies of the underlying random vector.

There is some work in the random matrix theory literature similar to (8). Suppose one has n iid copies of a p-dimensional random vector, forming a $p \times n$ data matrix $\boldsymbol{X}$. Let $\hat{r}_{ij}$, $1 \leq i, j \leq p$, be the sample correlations. Jiang (2004) showed that the limiting distribution of $\max_{1 \leq i < j \leq p} |\hat{r}_{ij}|$, after suitable normalization, is Gumbel provided that each column of $\boldsymbol{X}$ consists of p iid entries, each having finite moment of some order higher than 30, and p/n converges to some constant. His work was followed and improved by Zhou (2007) and Liu, Lin, and Shao (2008). In a recent article, Cai and Jiang (2011) extended those results in two ways: the dimension p could grow exponentially as the sample size n approaches infinity, under exponential moment conditions; and the test statistic $\max_{|i-j|>s_n} |\hat{r}_{ij}|$ converges to the Gumbel distribution if each column of $\boldsymbol{X}$ is Gaussian and is s_n-dependent. The latter generalization is important since it is one of few results that allow dependent entries. Their method is Poisson approximation, see for example Arratia, Goldstein, and Gordon (1989), depending heavily on the fact that for each sample correlation to be considered, the corresponding entries are independent. Schott (2005) proved that $\sum_{1 \leq i < j \leq p} \hat{r}_{ij}^2$ converges to a normal distribution after suitable normalization, under the conditions that each column of $\boldsymbol{X}$ contains iid Gaussian entries and p/n converges to some positive constant. His proof depends heavily on the normality assumption. Techniques developed in those papers are not applicable here since we have *only one realization* and the dependence structure among the entries can be quite complicated.

1.4. A Summary of Results of the Paper

Our main results are in Section 2, including a central limit theory of (2) and the Gumbel convergence (8). The proofs of the main results are given in Section 4. We also report on simulation study in Section 3, where we design a simulation-based block of blocks bootstrapping procedure that improves the finite-sample performance.

There is a supplementary file,that contains the proofs of some intermediate results used in Section 4, as well as proofs of other theorems and corollaries in Section 2. We establish a normal comparison principle in Section S.4 that is of independent interest. In Section S.5 we present a sufficient condition for summability of joint cumulants that is a commonly used assumption in time series analysis. Some auxiliary results are proved in Section S.6.

2. Main Results

To develop an asymptotic theory for time series, it is necessary to impose suitable measures of dependence and structural assumptions for the underlying process $\{X_i\}$. Here we adopt the framework of Wu (2005).

Assume that $\{X_i\}$ is a stationary causal process of the form

$$X_i = g(\cdots, \epsilon_{i-1}, \epsilon_i), \tag{9}$$

where $\epsilon_i, i \in \mathbb{Z}$, are iid random variables, and g is a measurable function for which X_i is a properly defined random variable. We define an operator Ω_k as follows. Suppose $X = h(\epsilon_j, \epsilon_{i-1}, \ldots)$ is a random variable which is a function of the innovations $\epsilon_l, l \le j$, then $\Omega_k(X) := h(\epsilon_j, \ldots, \epsilon_{k+1}, \epsilon'_k, \epsilon_{k-1}, \ldots)$, where $(\epsilon'_k)_{k\in\mathbb{Z}}$ is an iid copy of $(\epsilon_k)_{k\in\mathbb{Z}}$. Here ϵ_k in the definition of X is replaced by ϵ'_k.

For a random variable X and $p > 0$, we write $X \in \mathcal{L}^p$ if $\|X\|_p := (\mathbb{E}|X|^p)^{1/p} < \infty$. In particular, use $\|X\|$ for the $\mathcal{L}^2$-norm $\|X\|_2$. Assume $X_i \in \mathcal{L}^p$, $p > 1$. Define the *physical dependence measure of order* p as

$$\delta_p(i) = \|X_i - \Omega_0(X_i)\|_p, \tag{10}$$

which quantifying the dependence of X_i on the innovation ϵ_0. Our main results depend on the decay rate of $\delta_p(i)$ as $i \to \infty$. Let $p' = \min(2, p)$ and define

$$\begin{aligned} \Theta_p(n) &= \sum_{i=n}^{\infty} \delta_p(i), \quad \Psi_p(n) = \left(\sum_{i=n}^{\infty} \delta_p(i)^{p'} \right)^{1/p'}, \\ \Delta_p(n) &= \sum_{i=0}^{\infty} \min\{\mathcal{C}_p \Psi_p(n), \delta_p(i)\}, \end{aligned} \tag{11}$$

where $\mathcal{C}_p$ is $(p-1)^{-1}$ when $1 < p < 2$, and $\sqrt{p-1}$ when $p \ge 2$. It is easily seen that $\Psi_p(\cdot) \le \Theta_p(\cdot) \le \Delta_p(\cdot)$. We use Θ_p, Ψ_p, and Δ_p as shorthands for $\Theta_p(0)$, $\Psi_p(0)$ and $\Delta_p(0)$ respectively. We make the convention that $\delta_p(k) = 0$ for $k < 0$.

There are reasons to use the framework (9) and the dependence measure (10). First, the class of processes (9) is large, including linear processes, bilinear processes, Volterra processes, and many other time series models. See for instance Tong (1990) and Wu (2011). The physical dependence measure is easy to work with and is directly related to the underlying data-generating mechanism. The framework allows the development of an asymptotic theory for complicated statistics of time series.

2.1. Maximum deviations of sample autocovariances

Note that $\hat\gamma_k$ is a biased estimate of γ_k with $\mathbb{E}\hat\gamma_k = (1 - |k|/n)\gamma_k$. It is convenient to consider the centered version $\max_{1\le k\le s_n} \sqrt{n}|\hat\gamma_k - \mathbb{E}\hat\gamma_k|$ instead of $\max_{1\le k\le s_n} \sqrt{n}|\hat\gamma_k - \gamma_k|$. Recall (7) for a_n and b_n.

Theorem 1. *Assume $\mathbb{E}X_i = 0$, $X_i \in \mathcal{L}^p$ for some $p > 4$, and $\Theta_p(m) = O(m^{-\alpha})$, $\Delta_p(m) = O(m^{-\alpha'})$ for some $\alpha \ge \alpha' > 0$. If s_n satisfies $s_n \to \infty$ and $s_n = O(n^\eta)$ with*

$$0 < \eta < 1, \quad \eta < \alpha p/2, \ \ \textit{and} \ \ \eta \min\{2(p-2-\alpha p), (1-2\alpha')p\} < p-4, \tag{12}$$

then for all $x \in \mathbb{R}$,

$$\lim_{n\to\infty} P\left(\max_{1\le k\le s_n} |\sqrt{n}[\hat\gamma_k - (1-k/n)\gamma_k]| \le \sqrt{\sigma_0}(a_{2s_n}\, x + b_{2s_n}) \right) = \exp\{-\exp(-x)\}. \tag{13}$$

In (12), if $p \le 2 + \alpha p$ or $1 \le 2\alpha'$, then the second and third conditions are automatically satisfied, and hence Theorem 1 allows a very wide range of lags $s_n = O(n^\eta)$ with $0 < \eta < 1$. In this sense Theorem 1 is nearly optimal.

For the maximum deviation $\max_{1\le k<n} |\hat{\gamma}_k - \mathbb{E}\hat{\gamma}_k|$ over the range $1 \le k < n$, it seems not possible to derive a limiting distribution by using our method. However, we can obtain a sharp bound $(n^{-1}\log n)^{1/2}$. The upper bound is given in (15), while the lower bound can be obtained by applying Theorem 1 and choosing a sufficiently small η such that (12) holds. Using Theorem 2, Xiao and Wu (2012) derived convergence rates for the thresholded autocovariance matrix estimates.

Theorem 2. *Assume* $\mathbb{E}X_i = 0$, $X_i \in \mathcal{L}^p$ *for some* $p > 4$, *and* $\Theta_p(m) = O(m^{-\alpha})$, $\Delta_p(m) = O(m^{-\alpha'})$ *for some* $\alpha \ge \alpha' > 0$. *If*

$$\alpha > 1/2 \quad \textit{or} \quad \alpha' p > 2 \tag{14}$$

then for $c_p = 6(p+4)\, e^{p/4}\, \kappa_4\, \Theta_4$,

$$\lim_{n\to\infty} P\left(\max_{1\le k<n} |\hat{\gamma}_k - \mathbb{E}\hat{\gamma}_k| \le c_p \sqrt{\frac{\log n}{n}} \right) = 1. \tag{15}$$

Since $\Theta_p(m) \ge \Psi_p(m)$, we can assume $\alpha \ge \alpha'$. For a detailed discussion of their relationship, see Remark 6 of Xiao and Wu (2012). It turns out that for the special case of linear processes (12) can be weakened to

$$0 < \eta < 1, \quad \eta < \alpha p/2, \quad \text{and} \quad (1-2\alpha)\eta < (p-4)/p. \tag{16}$$

See Remark S.1 of the supplement. Furthermore, for linear processes (14) can be relaxed to $\alpha p > 2$.

The mean $\mu = \mathbb{E}X_0$ is often unknown and we estimate it by the sample mean $\bar{X}_n = (1/n)\sum_{i=1}^n X_i$. The usual estimates of autocovariances and autocorrelations are

$$\breve{\gamma}_k = \frac{1}{n}\sum_{i=|k|+1}^{n} (X_{i-k} - \bar{X}_n)(X_i - \bar{X}_n) \quad \text{and} \quad \breve{r}_k = \breve{\gamma}_k/\breve{\gamma}_0,\ |k| \le n-1. \tag{17}$$

Corollary 3. *Theorem 1 and Theorem 2 still hold if we replace* $\hat{\gamma}_k$ *therein by* $\breve{\gamma}_k$. *Furthermore,*

$$\lim_{n\to\infty} P\left(\max_{1\le k\le s_n} \left|\sqrt{n}[\breve{r}_k - (1-k/n)r_k]\right| \le (\sqrt{\sigma_0}/\gamma_0)(a_{2s_n}\, x + b_{2s_n}) \right) = \exp\{-\exp(-x)\}.$$

Proof of Corollary 3. For the $\breve{\gamma}_k$ version of Theorem 1, it suffices to show that

$$\max_{1\le k\le s_n} \left|\sqrt{n}(\breve{\gamma}_k - \hat{\gamma}_k)\right| = o_P\left(\frac{1}{\sqrt{\log s_n}}\right). \tag{18}$$

Let $S_k = \sum_{i=1}^k X_i$. By Theorem 1 (iii) of Wu (2007), we have $\|\max_{1\le k\le n} |S_k|\| \le 2\sqrt{n}\Theta_2$. Since

$$\sum_{i=k+1}^{n} (X_{i-k} - \bar{X}_n)(X_i - \bar{X}_n) - \sum_{i=k+1}^{n} X_{i-k}X_i = -\bar{X}_n \sum_{i=1}^{n-k} X_i + \bar{X}_n \sum_{i=1}^{k} X_i - k\bar{X}_n^2,$$

we have (18). The proof of the $\breve{\gamma}_k$ version of Theorem 2 is similar. The assertion on sample autocorrelations can be proved easily, and details are omitted. $\square$

2.2. Box-Pierce tests

Box-Pierce tests (Box and Pierce (1970); Ljung and Box (1978)) are commonly used in detecting lack of fit of a particular time series model. After a correct model has been fitted to a set of observations, one would expect the residuals to be close to a sequence of iid random variables, and therefore one should perform some tests for serial correlations as model diagnostics. Suppose $\{X_i\}_{1\le i\le n}$ is an iid sequence, let $\hat{r}_k$ be its sample autocorrelations. Then the distribution of $Q_n(K) := n\sum_{k=1}^K \hat{r}_K^2$ is approximately χ_K^2. Logically, it is not sufficient to consider a fixed number of correlations as the number of observations increases, because there may be some dependence at large lags. We present a normal theory about the Box-Pierce test statistic that allows the number of correlations included in Q_n to go to infinity.

Theorem 4. *Assume $X_i \in \mathcal{L}^8$, $\mathbb{E}X_i = 0$ and $\sum_{k=0}^\infty k^6\delta_8(k) < \infty$. If $s_n \to \infty$ and $s_n = O(n^\beta)$ for some $\beta < 1$, then*

$$\frac{1}{\sqrt{s_n}}\sum_{k=1}^{s_n}\left[n(\hat{\gamma}_k - (1-k/n)\gamma_k)^2 - (1-k/n)\sigma_0\right] \Rightarrow \mathcal{N}\left(0, 2\sum_{k\in\mathbb{Z}}\sigma_k^2\right).$$

To see the connection to the Box-Pierce test, we have a result on autocorrelations. Using the same argument, we can show that the same asymptotic law holds for the Ljung-Box test statistic $Q_{LB} = n(n+2)\sum_{k=1}^K \hat{r}_K^2/(n-k)$.

Corollary 5. *Under the conditions of Theorem 4, the same result holds if $\hat{\gamma}_k$ is replaced by $\check{\gamma}_k$. Furthermore,*

$$\frac{1}{\sqrt{s_n}}\sum_{k=1}^{s_n}\left[n(\hat{r}_k - (1-k/n)r_k)^2 - (1-k/n)\sigma_0/\gamma_0^2\right] \Rightarrow \mathcal{N}\left(0, \frac{2}{\gamma_0^4}\sum_{k\in\mathbb{Z}}\sigma_k^2\right). \tag{19}$$

The main condition of Theorem 4 and Corollary 5 is on how strong the dependence is. For example, if $X_i = \sum_{j=0}^\infty a_j\epsilon_{i-j}$ is a linear process, where ϵ_i are i.i.d. with mean zero, then the condition $\sum_{k=1}^\infty k^6\delta_8(k) < \infty$ is satisfied provided $\sum_{k=1}^\infty k^6|a_k| < \infty$ and $\mathbb{E}\epsilon_i^8 < \infty$. In the previous work, Hong (1996) considered i.i.d. processes; Durlauf (1991) and Deo (2000) studied martingale differences and required finite eighth moment of the underlying process; Shao (2011) considered general white noise whose physical dependence measures decay exponentially fast.

Remark 1. Theorem 4 clarifies an issue in the test of correlations. If $\gamma_k = 0$ for all $k \ge 1$, which means X_i are uncorrelated, then $\sigma_0 = \gamma_0^2$ and $\sigma_k = 0$ for all $|k| \ge 1$, and (19) becomes

$$\frac{1}{\sqrt{s_n}}\sum_{k=1}^{s_n}\left[n\hat{r}_k^2 - (1-k/n)\right] \Rightarrow \mathcal{N}(0,2). \tag{20}$$

In an influential paper, Romano and Thombs (1996) argued that, for fixed K, the chi-squared approximation for $Q_n(K)$ does not hold if X_i are only uncorrelated but not independent. One of the main reasons is that for fixed K, $\hat{r}_1, \ldots, \hat{r}_K$ are not asymptotically independent if X_i are not independent. The situation is different if the number of correlations included in Q_n can increase to infinity. According to (5), $\sqrt{n}\hat{\gamma}_{k_n}$ and $\sqrt{n}\hat{\gamma}_{k_n+h}$ are asymptotically independent if $h > 0$ and $k_n \to \infty$, because the asymptotic covariance is $\sigma_h = 0$. Therefore, the original Box-Pierce approximation of $Q_n(s_n)$ by $\chi^2_{s_n}$, with unbounded s_n, is still asymptotically valid as in (20) since $(\chi^2_{s_n} - s_n)/\sqrt{s_n} \Rightarrow \mathcal{N}(0,2)$ as $s_n \to \infty$. For example, consider the model $X_i = Z_i Z_{i-1}$ used in the simulation study of Romano and Thombs (1996), where Z_i are i.i.d. standard normal. Simulation shows that $\chi^2_{s_n}$ is a reasonably good approximation of $\sum_{k=1}^{s_n} n\hat{r}_k^2$ when $n = 10^3$ and $s_n = 50$. This observation again suggests that the asymptotic behaviors for bounded and unbounded lags are different. A similar observation has been made in Shao (2011), whose result also suggests that (20) is true under the assumption that $\delta_8(k) = O(\rho^k)$ for some $0 < \rho < 1$. They also considered non-uniform weights of sample autocorrelations by using kernel functions. Our condition $\sum_{k=1}^{\infty} k^6 \delta_8(k) < \infty$ is much weaker.

By allowing large s_n, the test can be powerful for detecting weak but long-range persistent correlations. To illustrate, we idealize that the $\hat{r}_k$ are independent and normally distributed as $\hat{r}_k \sim N(\delta_k, 1/n)$, where the δ_k are small but with similar order of magnitude. Then the Box-Pierce tests with small s_n may fail to reject the null hypothesis $H_0 : \delta_1 = \ldots = 0$. However, for large s_n, if $s_n^{1/2} = o(n\sum_{k=1}^{s_n} \delta_k^2)$, then H_0 can be rejected. □

Our next result has two closely related parts, one is on the estimation of $\sigma_0 = \sum_{k\in\mathbb{Z}} \gamma_k^2$, and the other is related to the power of the Box-Pierce test. Define the projection operator

$$\mathcal{P}^j \cdot = \mathbb{E}(\cdot|\mathcal{F}_{-\infty}^j) - \mathbb{E}(\cdot|\mathcal{F}_{-\infty}^{j-1}), \text{ where } \mathcal{F}_i^j = \langle \epsilon_i, \epsilon_{i+1}, \ldots, \epsilon_j \rangle,\ i, j \in \mathbb{Z}.$$

Theorem 6. *Assume $X_i \in \mathcal{L}^4$, $\mathbb{E}X_i = 0$, and $\Theta_4 < \infty$. If $s_n \to \infty$ and $s_n = o(\sqrt{n})$, then*

$$\sqrt{n}\left(\sum_{k=-s_n}^{s_n} \hat{\gamma}_k^2 - \sum_{k=-s_n}^{s_n} \gamma_k^2\right) \Rightarrow \mathcal{N}(0, 4\|D_0'\|^2), \tag{21}$$

where $D_0' = \sum_{i=0}^{\infty} \mathcal{P}^0(X_i Y_i)$ with $Y_i = \gamma_0 X_i + 2\sum_{k=1}^{\infty} \gamma_k X_{i-k}$. Furthermore, if $\sum_{k=1}^{\infty} \gamma_k^2 > 0$, then

$$\sqrt{n}\left(\sum_{k=1}^{s_n} \hat{\gamma}_k^2 - \sum_{k=1}^{s_n} \gamma_k^2\right) \Rightarrow \mathcal{N}(0, 4\|D_0\|^2), \tag{22}$$

where $D_0 = \sum_{i=0}^{\infty} \mathcal{P}^0(X_i Y_i)$ with $Y_i = \sum_{k=1}^{\infty} \gamma_k X_{i-k}$.

Corollary 7. *Under the conditions of Theorem 6, the same results hold if $\hat{\gamma}_k$ is replaced by $\breve{\gamma}_k$. Furthermore, there exist positive numbers τ_1^2 and τ_2^2 such that*

$$\sqrt{n}\left(\sum_{k=1}^{s_n} \hat{r}_k^2 - \sum_{k=1}^{s_n} r_k^2\right) \Rightarrow \mathcal{N}(0, \tau_1^2) \quad \textit{and} \quad \sqrt{n}\left(\sum_{k=-s_n}^{s_n} \hat{r}_k^2 - \sum_{k=-s_n}^{s_n} r_k^2\right) \Rightarrow \mathcal{N}(0, \tau_2^2).$$

As an immediate application, we consider the power of testing whether $\{X_i\}$ is an uncorrelated sequence. Battaglia (1990) studied the power of portmanteau tests with bounded lags. Hong (1996) considered the consistency and the asymptotic local power. Shao (2011) also briefly discussed the local power. Fan (1996) studied the power in a different but closely related context. According to (20), we can use the test statistic with unbounded lags

$$T_n := \frac{1}{\sqrt{s_n}} \left[Q_n(s_n) - \frac{s_n(2n - s_n - 1)}{2n} \right],$$

whose asymptotic distribution under the null hypothesis is $\mathcal{N}(0, 2)$. The null is rejected when $T_n > \sqrt{2} z_{1-\alpha}$, where $z_{1-\alpha}$ is the $(1-\alpha)$-th quantile of a standard normal random variable Z. However, under the alternative hypothesis $\sum_{k=1}^{\infty} r_k^2 > 0$, the distribution of T_n is approximated according to Corollary 7, and has asymptotic power

$$P\left(T_n > \sqrt{2} z_{1-\alpha}\right) \approx P\left(\tau_1 Z > \frac{\sqrt{2s_n} \cdot z_{1-\alpha}}{\sqrt{n}} + \frac{s_n(2n - s_n - 1)}{2n^{3/2}} - \sqrt{n} \sum_{k=1}^{s_n} r_k^2 \right),$$

which increases to 1 as n goes to infinity.

3. Blocks of Blocks Bootstrapping

If $\left(r_k^{(0)}\right)$ is a sequence of autocorrelations, one might be interested in the hypothesis test that $r_k = r_k^{(0)}$ for all $k \geq 1$. This cannot be tested in practice, except in some special parametric cases. A more tractable hypothesis is

$$\mathbf{H}_0 : \quad r_k = r_k^{(0)} \quad \text{for } 1 \leq k \leq s_n. \tag{23}$$

In traditional asymptotic theory, one often assumes that s_n is a fixed constant, for example, the popular Box-Pierce test for serial correlation. Our results provide both $\mathcal{L}^\infty$ and $\mathcal{L}^2$-based tests that allow s_n to grow as n increases. Nonetheless, the asymptotic tests can perform poorly when the sample size n is not large enough, and there may exist noticeable differences between the true and nominal probabilities of rejecting $\mathbf{H}_0$ (hereafter referred as error in rejection probability or ERP). Horowitz *et al.* (2006) showed that the Box-Pierce test with bootstrap-based p-values can significantly reduce the ERP. They used the blocks of blocks bootstrapping with overlapping blocks (hereafter referred as BOB) of Künsch (1989) and Liu and Singh (1992). For finite samples, our $\mathcal{L}^2$-based test is similar to the traditional Box-Pierce test considered in their paper, so we focus on the $\mathcal{L}^\infty$-based tests. We provide simulation evidence showing that the BOB works reasonably well.

Throughout this section, we let the innovations ϵ_i be iid standard normal random variables, and consider four models:

$$\text{I.I.D.} \qquad X_i = \epsilon_i; \tag{24}$$

$$\text{AR(1)} \qquad X_i = bX_{i-1} + \epsilon_i; \tag{25}$$

$$\text{Bilinear} \qquad X_i = (a + b\epsilon_i)X_{i-1} + \epsilon_i; \tag{26}$$

$$\text{ARCH} \qquad X_i = \sqrt{a + bX_{i-1}^2} \cdot \epsilon_i. \tag{27}$$

We generated each process with length $n = 2 \times 10^3,\ 2 \times 10^4,\ 2 \times 10^5,\ 2 \times 10^7$, and computed

$$a_{2s_n}^{-1} \left(\max_{1 \le k \le s_n} \sqrt{n}\, |\hat{r}_k - (1 - k/n) r_k| \,/ \sqrt{\hat{\sigma}_0} - b_{2s_n} \right) \tag{28}$$

with $s_n = 200,\ 2 \times 10^3,\ 10^4,\ 5 \times 10^5$, and $\hat{\sigma}_0 = \sum_{k=-t_n}^{t_n} \hat{r}_k^2$, where $t_n = \lfloor n^{1/3} \rfloor$. Based on 1000 repetitions, we plot the empirical distribution functions in Figure 1. We see that when $n = 2 \times 10^7$ and $s_n = 5 \times 10^5$, the four thickest long–dashed empirical curves are close to that of the Gumbel distribution, which confirms our theoretical results.

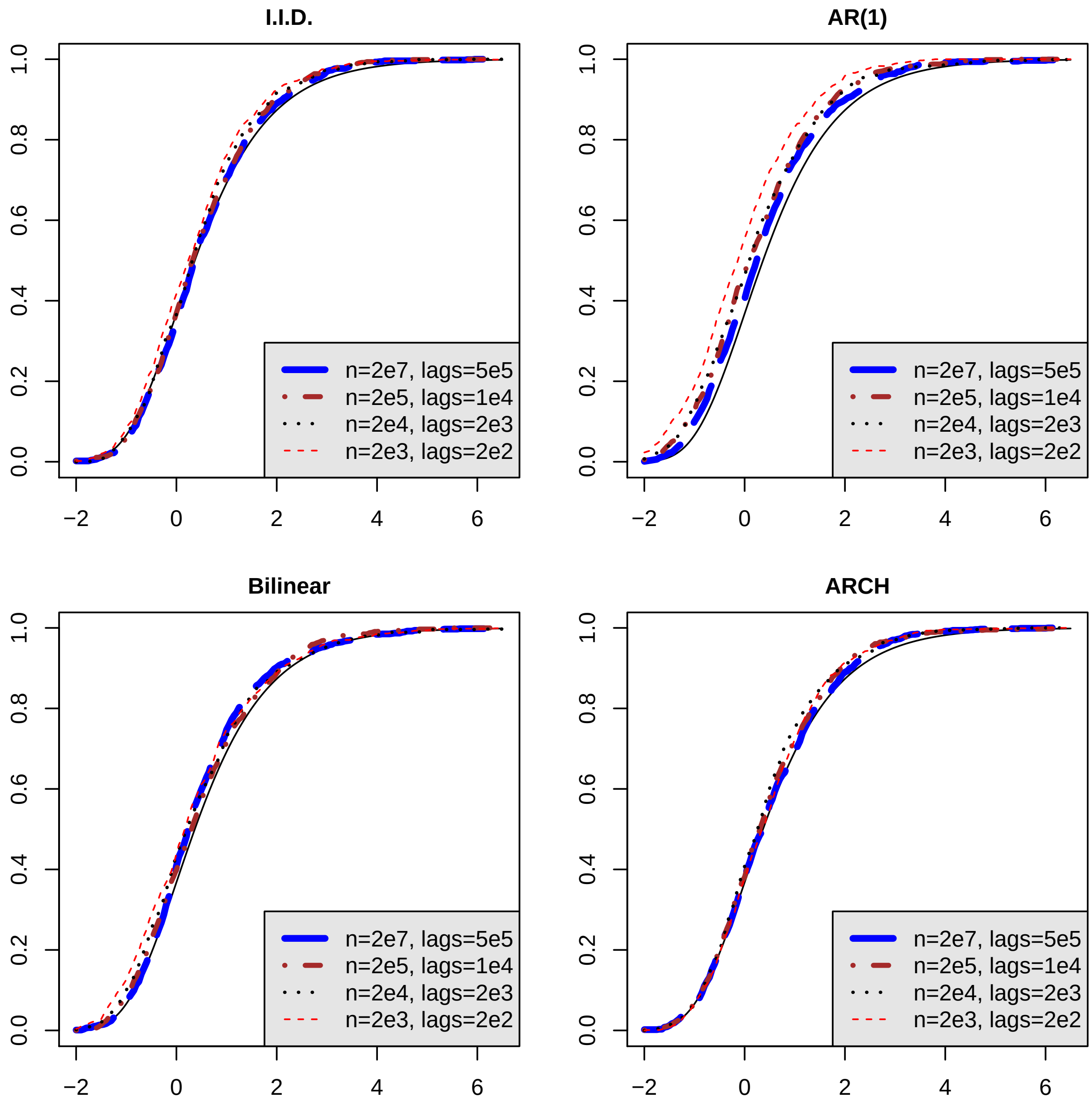

FIG 1. *Empirical distribution functions for quantities in (28). We chose $b = 0.5$ for model (25), $a = b = 0.4$ for model (26), and $a = b = 0.25$ for model (27). The solid line gives the true distribution function of the Gumbel distribution.*

On the other hand, the empirical distributions (yellow, green and red curves) are not very close to the limiting one if the sample sizes are not large-the Gumbel type of convergence in (13) is slow. This is a well-known phenomenon; see for example Hall (1979). It is therefore not reasonable to use the limiting distribution to approximate the finite sample distributions. To perform the test (23), we repeat the BOB procedure of Horowitz *et al.* (2006) (called SBOB in their paper). Since in the bootstrapped tests, the test statistics are not to be compared with the limiting distribution, we can ignore the norming constants in (28) and simply use the test statistics

$$M_n = \max_{1 \le k \le s_n} \left| \breve{r}_k - (1 - k/n) r_k^{(0)} \right| \quad \text{and} \quad \mathcal{M}_n = \frac{M_n}{\sqrt{\hat{\sigma}_0}} ,$$

where $\mathcal{M}_n$ is the self-normalized version with σ_0 estimated as $\hat{\sigma}_0 = \sum_{k=-t_n}^{t_n} \breve{r}_k^2$, and $t_n = \min\{\lfloor n^{1/3} \rfloor, s_n\}$. For simplicity, we refer to these tests as the M-test and the $\mathcal{M}$-test, respectively.

From the series $X_1, \ldots, X_n$, for some specified number of lags s_n included in the test and block size $\mathfrak{b}_n$, form $Y_i = (X_i, X_{i+1}, \ldots, X_{i+s_n})^\top$, $1 \le i \le n - s_n$ and blocks $\mathcal{B}_j = (Y_j, Y_{j+1}, \ldots, Y_{j+\mathfrak{b}_n - 1})$, $1 \le j \le n - s_n - \mathfrak{b}_n + 1$. For simplicity assume $h_n = n/\mathfrak{b}_n$ is an integer. Suppose $Y_\sharp$ is obtained by sampling a block $\mathcal{B}_\sharp$ from the set of blocks $\{\mathcal{B}_1, \mathcal{B}_2, \ldots, \mathcal{B}_{n-s_n-\mathfrak{b}_n+1}\}$, and then sampling a column from $\mathcal{B}_\sharp$, let $\mathrm{Cov}_\sharp$ represent the covariance of the bootstrap distribution of $Y_\sharp$, conditional on $(X_1, X_2, \ldots, X_n)$. Denote by $Y_\sharp^j$ the j-th entry of $Y_\sharp$, and set

$$r_k^{(e)} = \frac{\mathrm{Cov}_\sharp(Y_\sharp^1, Y_\sharp^{k+1})}{\sqrt{\mathrm{Cov}_\sharp(Y_\sharp^1, Y_\sharp^1) \cdot \mathrm{Cov}_\sharp(Y_\sharp^{k+1}, Y_\sharp^{k+1})}}.$$

The explicit formula for $r_k^{(e)}$ was given in Horowitz *et al.* (2006). The BOB algorithm is as follows.

1. Sample h_n times with replacement from $\{\mathcal{B}_1, \mathcal{B}_2, \ldots, \mathcal{B}_{n-s_n-\mathfrak{b}_n+1}\}$ to obtain blocks $\{\mathcal{B}_1^*, \mathcal{B}_2^*, \ldots, \mathcal{B}_{h_n}^*\}$ that are laid end-to-end to form a series of vectors $(Y_1^*, Y_2^*, \ldots, Y_n^*)$.
2. Take $(Y_1^*, Y_2^*, \ldots, Y_n^*)$ as a random sample of size n from some s_n-dimensional population distribution, and let r_k^* be the sample correlation of the first entry and the $(k+1)$-th entry. Calculate the test statistic $M_n^* = \max_{1 \le k \le s_n} \left| r_k^* - r_k^{(e)} \right|$ and $\mathcal{M}_n^* = M_n^* / \sqrt{\sigma_0^*}$, where $\sigma_0^* = \sum_{k=-t_n}^{t_n} (r_k^*)^2$.
3. Repeat steps 1 and 2 for N times. The bootstrap p-value of the M-test is given by $\#(M_n^* > M_n)/N$. For a nominal level α, we reject $\mathbf{H}_0$ if $\#(M_n^* > M_n)/N < \alpha$. The $\mathcal{M}$-test is performed in the same manner.

We compared the BOB tests and the asymptotic tests for the four models listed at the beginning of this section, with $a = .4$ for (25), $a = b = .4$ for (26) and $a = b = .25$ for (27). We set the series length as $n = 1800$, and considered four choices of s_n: $\lfloor \log(n) \rfloor = 7$, $\lfloor n^{1/3} \rfloor = 12$, $\lfloor \sqrt{n} \rfloor = 42$, and 25. The BOB tests were performed with $N = 999$, and the asymptotic tests were carried out by comparing $a_{2s_n}^{-1}(\sqrt{n}\mathcal{M}_n - b_{2s_n})$ with the corresponding quantiles of the Gumbel distribution. The empirical rejection probabilities based on 10,000 repetitions are reported in Table 1. All probabilities are given in percentages. For all cases, we see that the asymptotic tests are too conservative, and the ERP are quite large. At the nominal level 1%, the rejection probabilities are often around 0.1% or less, and are at most 0.51%; at nominal level 10%, they are often less than 3% and are at most 6.4%. Except for the bilinear models with $s_n = 7$ and $s_n = 12$, the bootstrapped tests significantly reduce the ERP: often less than 0.2% at nominal level 1%, less than .5% at level 5%, and less than 1% at level 10%. The performance of the M-test and the $\mathcal{M}$-test are similar, with the former being slightly more conservative. The BOB tests are relatively insensitive to the block size, which provides additional evidence of the findings on BOB tests in Davison and Hinkley (1997).

TABLE 1

Empirical rejection probabilities (in percentages)

Test	$s_n = 7$			$s_n = 12$			$s_n = 25$			$s_n = 42$		
	1	5	10	1	5	10	1	5	10	1	5	10
I.I.D.	.00	.34	1.6	.02	.69	2.3	.03	.93	3.2	.04	1.0	3.3
$\mathfrak{b}_n = 5$	1.3	5.1	10.0	1.1	5.2	9.8	.95	4.7	9.3	1.0	4.7	9.6
	1.4	5.3	10.4	1.2	5.6	10.5	1.1	5.1	10.1	1.1	5.1	10.2
$\mathfrak{b}_n = 10$	.83	4.8	10.0	1.1	4.9	9.6	1.1	4.9	10.1	.65	4.3	8.9
	.94	5.1	10.3	1.2	5.4	10.3	1.1	5.5	11.0	.78	4.7	9.6
AR(1)	.01	.17	1.2	.01	.36	1.8	.02	.77	2.5	.02	.88	2.8
$\mathfrak{b}_n = 10$	1.3	5.7	10.9	1.3	5.5	11.4	1.3	5.5	10.9	1.1	5.7	11.2
	1.3	5.7	11.2	1.4	5.9	11.7	1.3	6.0	11.5	1.2	6.0	11.7
$\mathfrak{b}_n = 20$	.98	5.5	10.9	1.0	5.8	11.3	1.1	5.3	10.6	.86	4.9	10.5
	1.0	5.7	11.0	1.1	6.1	11.9	1.2	5.6	11.0	.83	5.0	10.9
Bilinear	.34	2.8	6.4	.43	2.5	5.8	.51	2.5	5.9	.40	2.8	5.9
$\mathfrak{b}_n = 10$	2.8	8.7	14.4	1.8	7.1	12.7	1.2	6.1	12.0	1.2	5.4	10.9
	2.7	8.6	14.5	1.8	7.3	12.9	1.3	6.2	12.2	1.1	5.5	11.1
$\mathfrak{b}_n = 20$	2.7	8.4	14.6	2.1	7.2	13.5	1.5	6.3	12.0	1.3	5.2	10.8
	2.5	8.3	14.6	2.1	7.5	13.9	1.5	6.2	12.0	1.2	5.3	10.9
ARCH	.05	.82	3.2	.06	1.5	3.9	.09	1.3	4.0	.12	1.4	4.4
$\mathfrak{b}_n = 10$	.99	5.0	10.5	1.2	4.9	9.7	.80	4.6	9.9	.82	4.7	9.3
	1.1	5.4	10.9	1.4	5.3	10.4	.92	5.1	10.7	.94	5.1	10.2
$\mathfrak{b}_n = 20$	.86	5.1	10.5	1.0	5.0	10.3	.69	4.8	9.7	.63	4.3	8.9
	.98	5.5	11.0	1.2	5.6	11.0	.89	5.1	10.4	.76	4.7	9.5

The values 1, 5, 10 in the 2nd row indicate nominal levels in percentages. The numbers in the third row starting with the model name "I.I.D." are for the asymptotic tests. The fourth row starting with $\mathfrak{b}_n = 5$ is for BOB M-tests with block size 5. The fifth row is for BOB $\mathcal{M}$-tests with the same block size 5. Other rows should be read similarly.

The bootstrapped tests still perform relatively poorly for bilinear models when s_n is small ($s_n = 7$ and 12). This is possibly due to the heavy-tail of the bilinear process. Tong (1981) gave necessary conditions for the existence of even order moments. Horowitz *et al.* (2006) showed that the iterated bootstrapping further reduce the ERP. It is of interest to see whether the iterated procedure has the same effect for the $\mathcal{L}^\infty$ based test, in particular, whether it makes the ERP reasonably small for the bilinear models when s_n is small. The simulation for the iterated bootstrapping is computationally expensive and we do not pursue it here.

Remark 2. Assume the time series is governed by a parameter or a set of parameters ρ. Let $\hat\rho$ be an estimate of ρ, then the autocovariance estimates given by the parametric model are $\gamma_k(\hat\rho)$. The results of Theorem 1 and Theorem 4 can be used for diagnostic checking if we replace the true autocovariances γ_k by $\gamma_k(\hat\rho)$. The same asymptotic results hold for a broad class of parametric models. For example, consider the AR(1) model $X_i = \rho X_{i-1} + \epsilon_i$, where $|\rho| < 1$. Without loss of generality, take $\mathbb{E}X_i = 0$ and $\mathrm{Var}(X_i) = 1$. Let $\hat\rho$ be a $\sqrt{n}$

consistent estimate of ρ. By Lemma 11 and

$$\max_{t_n \le k \le s_n} |(1 - k/n)\sqrt{n} \cdot (\hat{\rho}^k - \rho^k)| = O_p\left[(\rho/2 + 1/2)^{t_n}\right],$$

the limiting distribution in (13) remains true if we replace γ_k by $\hat{\rho}^k$. On the other hand, we also have

$$\sum_{k=t_n}^{s_n} n(\hat{\rho}^k - \rho^k)^2 = O_p\left[(\rho/2 + 1/2)^{t_n}\right]$$

if the sequence (t_n) satisfies $t_n \to \infty$ and $t_n = o(s_n)$; hence Theorem 4 still holds if we replace γ_k by $\hat{\rho}^k$. We will consider the impact of plug-in estimates under a more general setting in future research.

4. Proofs

In this section we prove Theorem 1 and Theorem 4. Since the proofs are lengthy, we only provide the major steps and ideas, and leave most technical details to a supplementary file. The proofs of other theorems and corollaries are also given in the supplementary file.

We list some notations here. The operator $\mathbb{E}_0$ is defined as $\mathbb{E}_0 X := X - \mathbb{E}X$ for any random variable X. For a vector $\boldsymbol{x} = (x_1, \ldots, x_d)^\top \in \mathbb{R}^d$, let $|\boldsymbol{x}|$ be the Euclidean norm, $|\boldsymbol{x}|_\infty := \max_{1 \le i \le d} |x_i|$, and $|\boldsymbol{x}|_\bullet := \min_{1 \le i \le d} |x_i|$. We use C to denote a constant whose values may vary from place to place. A constant with a symbolic subscript is used to emphasize the dependence of the value on the subscript.

The framework (9) is particularly suited for two classical tools for dealing with dependent sequences, martingale approximation and m-dependence approximation. For $i \le j$, let $\mathcal{F}_i^j = \langle \epsilon_i, \epsilon_{i+1}, \ldots, \epsilon_j \rangle$ be the σ-field generated by the innovations $\epsilon_i, \epsilon_{i+1}, \ldots, \epsilon_j$, and define the projection operator $\mathcal{H}_i^j(\cdot) = \mathbb{E}(\cdot|\mathcal{F}_i^j)$. Set $\mathcal{F}_i := \mathcal{F}_i^\infty$, $\mathcal{F}^j := \mathcal{F}_{-\infty}^j$, and define $\mathcal{H}_i$ and $\mathcal{H}^j$ similarly. Given the projection operators $\mathcal{P}^j(\cdot) = \mathcal{H}^j(\cdot) - \mathcal{H}^{j-1}(\cdot)$, and $\mathcal{P}_i(\cdot) = \mathcal{H}_i(\cdot) - \mathcal{H}_{i+1}(\cdot)$, $(\mathcal{P}^j(\cdot))_{j \in \mathbb{Z}}$ and $(\mathcal{P}_{-i}(\cdot))_{i \in \mathbb{Z}}$ are martingale difference sequences with respect to the filtrations $(\mathcal{F}^j)$ and $(\mathcal{F}_{-i})$, respectively. For $m \ge 0$, take $\tilde{X}_i = \mathcal{H}_{i-m} X_i$, then $(\tilde{X}_i)_{i \in \mathbb{Z}}$ is a $(m+1)$-dependent sequence.

4.1. Proof of Theorem 1

We give an outline of intermediate steps, then conclude with the proof of Theorem 1. The proofs of intermediate lemmas are provided in Section S.2 of the supplementary file, as are the proofs of other results in Section 2.1.

Step 1: m-dependence approximation. Define $R_{n,k} = \sum_{i=k+1}^n (X_{i-k} X_i - \gamma_k)$. Set $m_n = \lfloor n^\beta \rfloor$, $0 < \beta < 1$. Take $\tilde{X}_i = \mathcal{H}_{i-m_n} X_i$, $\tilde{\gamma}_k = \mathbb{E}(\tilde{X}_0 \tilde{X}_k)$, and $\tilde{R}_{n,k} = \sum_{i=k+1}^n (\tilde{X}_{i-k} \tilde{X}_i - \tilde{\gamma}_k)$.

Lemma 8. *Assume $\mathbb{E}X_i = 0$, $X_i \in \mathcal{L}^p$, and $\Theta_p(m) = O(m^{-\alpha})$ for some $p > 4$ and $\alpha > 0$. If $s_n = O(n^\eta)$ with $0 < \eta < \alpha p/2$, then there exists a β such that $\eta < \beta < 1$ and*

$$\max_{1 \le k \le s_n} \left| R_{n,k} - \tilde{R}_{n,k} \right| = o_P\left(\sqrt{n/\log s_n}\right).$$

Step 2: Throw out small blocks. Let $l_n = \lfloor n^\gamma \rfloor$, where $\gamma \in (\beta, 1)$. For each $t_n < k \le s_n$, we split the integer interval $[k+1, n]$ into alternating large and small blocks

$$
\begin{aligned}
&K_1 = [k+1, s_n] \\
&H_j = [s_n + (j-1)(2m_n + l_n) + 1, s_n + (j-1)(2m_n + l_n) + l_n]; \quad 1 \le j \le w_n - 1, \\
&K_{j+1} = [s_n + (j-1)(2m_n + l_n) + l_n + 1, s_n + j(2m_n + l_n)]; \quad 1 \le j \le w_n - 1; \quad \text{and} \\
&H_{w_n} = [s_n + (w_n - 1)(2m_n + l_n) + 1, n],
\end{aligned}
\tag{29}
$$

where w_n is the largest integer such that $s_n + (w_n - 1)(2m_n + l_n) + l_n \le n$. Denote by $|H|$ the size of a block H. By definition, $l_n \le |H_{w_n}| \le 3 l_n$ when n is large enough. For $1 \le j \le w_n$, define

$$
V_{k,j} = \sum_{i \in K_j,\, i > k} \left(\tilde{X}_{i-k} \tilde{X}_i - \tilde{\gamma}_k \right) \text{ and } U_{k,j} = \sum_{i \in H_j} \left(\tilde{X}_{i-k} \tilde{X}_i - \tilde{\gamma}_k \right).
$$

Note that $w_n \sim n/(2m_n + l_n) \sim n^{1-\gamma}$.

Lemma 9. *Under the conditions of Theorem 1,*

$$
\max_{1 \le k \le s_n} \left| \sum_{j=1}^{w_n} V_{k,j} \right| = o_P\left(\sqrt{\frac{n}{\log s_n}} \right).
$$

Step 3: Truncate sums over large blocks. We show that it suffices to consider

$$
\mathcal{R}_{n,k} = \sum_{j=1}^{w_n} \bar{U}_{k,j}, \text{ where } \bar{U}_{k,j} = \mathbb{E}_0 \left(U_{k,j} I\{|U_{k,j}| \le \sqrt{n}/(\log s_n)^3\} \right).
$$

where $I\{\cdot\}$ is the indicator function.

Lemma 10. *Under the conditions of Theorem 1,*

$$
\max_{1 \le k \le s_n} \left| \sum_{j=1}^{w_n} (U_{k,j} - \bar{U}_{k,j}) \right| = o_P\left(\sqrt{\frac{n}{\log s_n}} \right).
$$

Step 4: Compare covariance structures. In order to prove Lemma 13, we need the autocovariance structure of $(\mathcal{R}_{n,k}/\sqrt{n})$ to be close to that of (G_k). However, this only happens when k is large. We show that there exists an $0 < \iota < 1$ such that for $t_n = 3\lfloor s_n^\iota \rfloor$: $\max_{1 \le k \le t_n} |\mathcal{R}_{n,k}/\sqrt{n}|$ does not contribute to the asymptotic distribution; and the autocovariance structure of $(\mathcal{R}_{n,k}/\sqrt{n})$ converges to that of (G_k) uniformly on $t_n < k \le s_n$.

Lemma 11. *Under conditions of Theorem 1, there exists a constant $0 < \iota < 1$ such that for $t_n = 3\lfloor s_n^\iota \rfloor$,*

$$
\lim_{n \to \infty} P\left(\max_{1 \le k \le t_n} |\mathcal{R}_{n,k}| > \sqrt{\sigma_0 n \log s_n} \right) = 0. \tag{30}
$$

Lemma 12. *Under the conditions of Theorem 1, and with $t_n = 3\lfloor s_n^\iota \rfloor$, there exist a constant $C_p > 0$ and $0 < \ell < 1$ such that, for any $t_n < k \le k + h \le s_n$,*

$$
|\operatorname{Cov}(\mathcal{R}_{n,k}, \mathcal{R}_{n,k+h})/n - \sigma_h| \le C_p\, s_n^{-\ell}.
$$

Step 5: Moderate deviations. Let $t_n = 3\lfloor s_n^\iota \rfloor$ be as in Lemma 11. For $t_n < k_1 < k_2 < \ldots < k_d \leq s_n$, take $\boldsymbol{\mathcal{R}}_n = (\mathcal{R}_{n,k_1}, \mathcal{R}_{n,k_2}, \ldots, \mathcal{R}_{n,k_d})^\top$ and $\boldsymbol{V} = (G_{k_1}, G_{k_2}, \ldots, G_{k_d})^\top$, where (G_k) is defined in (6). Let $\Sigma_n = \mathrm{Cov}(\boldsymbol{\mathcal{R}}_n)$ and $\Sigma = \mathrm{Cov}(\boldsymbol{V})$. For fixed $x \in \mathbb{R}$, set $z_n = a_{2s_n}x + b_{2s_n}$, where the constants a_n and b_n are from (7).

Lemma 13. *Under conditions of Theorem 1, there exists a constant $C_{p,d} > 1$ such that, for all $t_n < k_1 < k_2 < \ldots < k_d \leq s_n$,*

$$\left|P\left(\left|\boldsymbol{\mathcal{R}}_n/\sqrt{n}\right|_\bullet \geq z_n\right) - P\left(|\boldsymbol{V}|_\bullet \geq z_n\right)\right| \leq C_{p,d}\frac{P\left(|\boldsymbol{V}|_\bullet \geq z_n\right)}{(\log s_n)^{1/2}} + C_{p,d}\exp\left\{-\frac{(\log s_n)^2}{C_{p,d}}\right\}.$$

We need a result on the Gaussian process that might be of independent interest.

Theorem 14. *Let (X_n) be a stationary mean zero Gaussian process, with $r_k = \mathrm{Cov}(X_0, X_k)$. Assume $r_0 = 1$, and $\lim_{n\to\infty} r_n(\log n) = 0$. If $a_n = (2\log n)^{-1/2}$, $b_n = (2\log n)^{1/2} - (8\log n)^{-1/2}(\log\log n + \log 4\pi)$, and $z_n = a_n z + b_n$ for $z \in \mathbb{R}$, with $A_i = \{X_i \geq z_n\}$ and*

$$Q_{n,d} = \sum_{1\leq i_1<\ldots<i_d\leq n} P(A_{i_1} \cap \cdots \cap A_{i_d}),$$

it holds that $\lim_{n\to\infty} Q_{n,d} = e^{-dz}/d\,!$ for all $d \geq 1$. Furthermore, the same result holds if $A_i = \{|X_i| \geq z_{2n}\}$.

The proofs of the preceding results are given in a supplementary file.

Proof of Theorem 1. Set $z_n = a_{2s_n}x + b_{2s_n}$. It suffices to show

$$\lim_{n\to\infty} P\left(\max_{t_n<k\leq s_n} |\mathcal{R}_k/\sqrt{n}| \leq \sqrt{\sigma_0}z_n\right) = \exp\{-\exp(-x)\}. \tag{31}$$

Without loss of generality assume $\sigma_0 = 1$. Take $A_k = \{G_k \geq z_n\}$ and $B_k = \{\mathcal{R}_k/\sqrt{n} \geq z_n\}$. Let

$$Q_{n,d} = \sum_{t_n<k_1<\ldots<k_d\leq s_n} P(A_{k_1} \cap \cdots \cap A_{k_d}) \quad \text{and} \quad \tilde{Q}_{n,d} = \sum_{t_n<k_1<\ldots<k_d\leq s_n} P(B_{k_1} \cap \cdots \cap B_{k_d}).$$

By the inclusion-exclusion formula, for any $q \geq 1$

$$\sum_{d=1}^{2q}(-1)^{d-1}\tilde{Q}_{n,d} \leq P\left(\max_{t_n<k\leq s_n} |\mathcal{R}_k/\sqrt{n}| \geq a_{2s_n}x + b_{2s_n}\right) \leq \sum_{d=1}^{2q-1}(-1)^{d-1}\tilde{Q}_{n,d}. \tag{32}$$

By Lemma 13, $|\tilde{Q}_{n,d} - Q_{n,d}| \leq C_{p,d}(\log s_n)^{-1/2}Q_{n,d} + s_n^{-1}$. By Theorem 14 with elementary calculations, $\lim_{n\to\infty} Q_{n,d} = e^{-dx}/d!$, and hence $\lim_{n\to\infty} \tilde{Q}_{n,d} = e^{-dx}/d!$. By letting n go to infinity and then d go to infinity in (32), we obtain (31), and the proof is complete. □

4.2. Proof of Theorem 4

We outline intermediate steps, and then prove Theorem 4. The proofs of intermediate lemmas and other results of Section 2.2 are given in Section S.3 of the supplementary file.

Step 1: m-dependence approximation. Without loss of generality, assume $s_n \le \lfloor n^\beta \rfloor$. Set $m_n = 2\lfloor n^\beta \rfloor$. Let $\tilde{X}_i = \mathcal{H}_{i-m_n}^i X_i$ and $\tilde{R}_{n,k} = \sum_{i=k+1}^n (\tilde{X}_{i-k}\tilde{X}_i - \tilde{\gamma}_k)$. By (S.7) and (S.13), if $\Theta_4(m) = o(m^{-\alpha})$ for some $\alpha > 0$, then for all $1 \le k \le s_n$,

$$\mathbb{E}|R_{n,k}^2 - \tilde{R}_{n,k}^2| \le \|R_{n,k} + \tilde{R}_{n,k}\| \cdot \|R_{n,k} - \tilde{R}_{n,k}\| \le C\,\Theta_4^3 \cdot n \cdot \Theta_4\left(m_n/2\right) = o\left(n^{1-\alpha\beta}\right).$$

The condition $\sum_{k=0}^\infty k^6 \delta_8(k) < \infty$ implies that $\Theta_4(m) = O(m^{-6})$. Therefore, under the conditions of Theorem 4, we have

$$\frac{1}{n\sqrt{s_n}} \sum_{k=1}^{s_n} \mathbb{E}_0\left(R_{n,k}^2 - \tilde{R}_{n,k}^2\right) = o_P(1).$$

Step 2: Throw out small blocks. Let $l_n = \lfloor n^\eta \rfloor$, where $\eta \in (\beta, 1)$. Split the interval $[1, n]$ into the blocks

$$\begin{aligned}
&K_0 = [1, s_n],\\
&H_j = [s_n + (j-1)(2m_n + l_n) + 1, s_n + (j-1)(2m_n + l_n) + l_n], \quad 1 \le j \le w_n,\\
&K_j = [s_n + (j-1)(2m_n + l_n) + l_n + 1, s_n + j(2m_n + l_n)], \quad 1 \le j \le w_n - 1, \quad \text{and}\\
&K_{w_n} = [s_n + (w_n - 1)(2m_n + l_n) + l_n + 1, n],
\end{aligned}$$

where w_n is the largest integer such that $s_n + (w_n - 1)(2m_n + l_n) + l_n \le n$. Take $U_{k,0} = 0$, $V_{k,0} = \sum_{i \in K_0, i > k} (\tilde{X}_{i-k}\tilde{X}_i - \tilde{\gamma}_k)$, and $U_{k,j} = \sum_{i \in H_j} (\tilde{X}_{i-k}\tilde{X}_i - \tilde{\gamma}_k)$, $V_{k,j} = \sum_{i \in K_j} (\tilde{X}_{i-k}\tilde{X}_i - \tilde{\gamma}_k)$ for $1 \le j \le w_n$. Set $\mathcal{R}_{n,k} = \sum_{j=1}^{w_n} U_{k,j}$. Observe that by construction, $U_{k,j}, 1 \le j \le w_n$ are iid random variables.

Lemma 15. *If $X_i \in \mathcal{L}^8$, $\mathbb{E}X_i = 0$, and $\sum_{k=0}^\infty k^6 \delta_8(k) < \infty$, then*

$$\frac{1}{n\sqrt{s_n}} \sum_{k=1}^{s_n} \mathbb{E}_0\left(\tilde{R}_{n,k}^2 - \mathcal{R}_{n,k}^2\right) = o_P(1).$$

Step 3: Central limit theorem concerning $\mathcal{R}_{n,k}$'s.

Lemma 16. *If $X_i \in \mathcal{L}^8$, $\mathbb{E}X_i = 0$, and $\sum_{k=0}^\infty k^6 \delta_8(k) < \infty$, then*

$$\frac{1}{n\sqrt{s_n}} \sum_{k=1}^{s_n} \left(\mathcal{R}_{n,k}^2 - \mathbb{E}\mathcal{R}_{n,k}^2\right) \Rightarrow \mathcal{N}\left(0, 2\sum_{k \in \mathbb{Z}} \sigma_k^2\right).$$

We are now ready to prove Theorem 4.

Proof of Theorem 4. By Lemma 15 and Lemma 16,

$$\frac{1}{n\sqrt{s_n}} \sum_{k=1}^{s_n} \left(R_{n,k}^2 - \mathbb{E}R_{n,k}^2\right) \Rightarrow \mathcal{N}\left(0, 2\sum_{k \in \mathbb{Z}} \sigma_k^2\right).$$

It remains to show that

$$\lim_{n\to\infty} \frac{1}{n\sqrt{s_n}} \sum_{k=1}^{s_n} \left[\mathbb{E}R_{n,k}^2 - (n-k)\sigma_0\right] = 0. \tag{33}$$

We need Lemma S.2 of the supplementary file with a slight modification. Observe that in equation (S.29) of the supplementary file, we have $\sum_{j=1}^{m_n} \Theta_2(j)^2 < \infty$, and hence

$$\left|\mathbb{E}R_{n,k}^2 - (n-k)\sigma_0\right| \leq C\left[(n-k)\Delta_4(\lfloor k/3\rfloor + 1) + \sqrt{n-k}\right].$$

With the condition $\Theta_8(m) = o(m^{-6})$, elementary calculations show that $\Delta_4(m) = o(m^{-5})$, hence (33) follows. The proof is complete. □

Acknowledgements. We are grateful to the Editor, an Associate Editor and the referees for their many helpful comments. Han Xiao's research was supported in part by the US National Science Foundation (DMS–1209091). Wei Biao Wu's research was supported in part by the US National Science Foundation (DMS–0906073 and DMS–1106970).

References

An, H. Z., Chen, Z. G., and Hannan, E. J. (1982). Autocorrelation, autoregression and autoregressive approximation. *Ann. Statist.* **10**, 926–936.

Anderson, T. W. (1991). The asymptotic distributions of autoregressive coefficients Technical Report No. 26, Stanford University, Department of Statistics.

Arratia, R., Goldstein, L., and Gordon, L. (1989). Two moments suffice for Poisson approximations: the Chen-Stein method. *Ann. Probab.* **17**, 9–25.

Battaglia, F. (1990). Approximate power of portmanteau tests for time series. *Statist. Probab. Lett.* **9**, 337–341.

Berman, S. M. (1964). Limit theorems for the maximum term in stationary sequences. *Ann. Math. Statist.* **35**, 502–516.

Bickel, P. J. and Levina, E. (2008a). Covariance regularization by thresholding. *Ann. Statist.* **36**, 2577–2604.

Bickel, P. J. and Levina, E. (2008b). Regularized estimation of large covariance matrices. *Ann. Statist.* **36**, 199–227.

Box, G. E. P. and Pierce, D. A. (1970). Distribution of residual autocorrelations in autoregressive-integrated moving average time series models. *J. Amer. Statist. Assoc.* **65**, 1509–1526.

Brockwell, P. J. and Davis, R. A. (1991). *Time Series: Theory and Methods*, Second ed. Springer-Verlag, New York.

Burkholder, D. L. (1988). Sharp inequalities for martingales and stochastic integrals. *Astérisque* 157-158, 75–94.

Cai, T. T. and Jiang, T. (2011). Limiting laws of coherence of random matrices with applications to testing covariance structure and construction of compressed sensing matrices. *Ann. Statist.* **39**, 1496–1525.

Davison, A. C. and Hinkley, D. V. (1997). *Bootstrap Methods and Their Application.* Cambridge University Press, Cambridge.

Deo, R. S. (2000). Spectral tests of the martingale hypothesis under conditional heteroscedasticity. *J. Econometrics* **99**, 291–315.

Duchesne, P., Li, L., and Vandermeerschen, J. (2010). On testing for serial correlation of unknown form using wavelet thresholding. *Computational Statistics and Data Analysis* **54**, 2512 - 2531.

Durbin, J. and Watson, G. S. (1950). Testing for serial correlation in least squares regression. I. *Biometrika* **37**, 409–428.

Durbin, J. and Watson, G. S. (1951). Testing for serial correlation in least squares regression. II. *Biometrika* **38**, 159–178.

Durlauf, S. N. (1991). Spectral based testing of the martingale hypothesis. *J. Econometrics* **50**, 355–376.

Escanciano, J. C. and Lobato, I. N. (2009). An automatic Portmanteau test for serial correlation. *J. Econometrics* **151**, 140–149.

Fan, J. (1996). Test of significance based on wavelet thresholding and Neyman's truncation. *J. Amer. Statist. Assoc.* **91**, 674–688.

Hall, P. (1979). On the rate of convergence of normal extremes. *J. Appl. Probab.* **16**, 433–439.

Hannan, E. J. (1974). The uniform convergence of autocovariances. *Ann. Statist.* **2**, 803–806.

Hannan, E. J. (1976). The asymptotic distribution of serial covariances. *Ann. Statist.* **4**, 396–399.

HANNAN, E. J. and DEISTLER, M. (1988). *The Statistical Theory of Linear Systems.* John Wiley & Sons Inc., New York.

HANNAN, E. J. and HEYDE, C. C. (1972). On limit theorems for quadratic functions of discrete time series. *Ann. Math. Statist.* **43**, 2058–2066.

HONG, Y. (1996). Consistent testing for serial correlation of unknown form. *Econometrica* **64**, 837–864.

HONG, Y. and LEE, Y. J. (2003). Consistent testing for serial uncorrelation of unknown form under general conditional heteroscedasticity. Preprint, Cornell University, Department of Economics.

HOROWITZ, J. L., LOBATO, I. N., NANKERVIS, J. C., and SAVIN, N. E. (2006). Bootstrapping the Box-Pierce Q test: A robust test of uncorrelatedness. *J. Econometrics* **133**, 841-862.

HOSOYA, Y. and TANIGUCHI, M. (1982). A central limit theorem for stationary processes and the parameter estimation of linear processes. *Ann. Statist.* **10**, 132–153.

JIANG, T. (2004). The asymptotic distributions of the largest entries of sample correlation matrices. *Ann. Appl. Probab.* **14**, 865–880.

JIRAK, M. (2011). On the maximum of covariance estimators. *Journal of Multivariate Analysis* **102**, 1032 - 1046.

KÜNSCH, H. R. (1989). The jackknife and the bootstrap for general stationary observations. *Ann. Statist.* **17**, 1217–1241.

LEE, J. and HONG, Y. (2001). Testing for serial correlation of unknown form using wavelet methods. *Econometric Theory* **17**, 386–423.

LI, W. K. (2003). *Diagnostic Checks in Time Series.* Chapman and Hall/CRC.

LIU, W.-D., LIN, Z., and SHAO, Q.-M. (2008). The asymptotic distribution and Berry-Esseen bound of a new test for independence in high dimension with an application to stochastic optimization. *Ann. Appl. Probab.* **18**, 2337–2366.

LIU, R. Y. and SINGH, K. (1992). Moving blocks jackknife and bootstrap capture weak dependence. In *Exploring the Limits of Bootstrap* 225–248. Wiley, New York.

LJUNG, G. and BOX, G. E. P. (1978). Measure of lack of fit in time-series models. *Biometrika* **65**, 297-303.

PAPARODITIS, E. (2000). Spectral density based goodness-of-fit tests for time series models. *Scand. J. Statist.* **27**, 143–176.

PAPARODITIS, E. and POLITIS, D. N. (2012). Nonlinear spectral density estimation: thresholding the correlogram. *J. Time Series Anal.* **33**, 386–397.

PHILLIPS, P. C. B. and SOLO, V. (1992). Asymptotics for linear processes. *Ann. Statist.* **20**, 971–1001.

POLITIS, D. N. (2003). Adaptive bandwidth choice. *J. Nonparametr. Stat.* **15**, 517–533.

ROBINSON, P. M. (1991). Testing for strong serial correlation and dynamic conditional heteroskedasticity in multiple regression. *J. Econometrics* **47**, 67–84.

ROMANO, J. P. and THOMBS, L. A. (1996). Inference for autocorrelations under weak assumptions. *J. Amer. Statist. Assoc.* **91**, 590–600.

SCHOTT, J. R. (2005). Testing for complete independence in high dimensions. *Biometrika* **92**, 951–956.

SHAO, X. (2011). Testing for white noise under unknown dependence and its applications to diagnostic checking for time series models. *Econometric Theory* **27**, 312–343.

TONG, H. (1981). A note on a Markov bilinear stochastic process in discrete time. *J. Time Ser. Anal.* **2**, 279–284.

TONG, H. (1990). *Nonlinear Time Series.* The Clarendon Press, Oxford University Press, New York.

WU, W. B. (2005). Nonlinear system theory: another look at dependence. *Proc. Natl. Acad. Sci. USA* **102**, 14150–14154.

WU, W. B. (2007). Strong invariance principles for dependent random variables. *Ann. Probab.* **35**, 2294–2320.

WU, W. B. (2009). An asymptotic theory for sample covariances of Bernoulli shifts. *Stochastic Process. Appl.* **119**, 453–467.

WU, W. B. (2011). Asymptotic theory for stationary processes. *Statistics and Its Interface* **4**, 207-226.

XIAO, H. and WU, W. B. (2012). Covariance matrix estimation for stationary time series. *Ann. Statist.* **40**, 466–493.

ZHOU, W. (2007). Asymptotic distribution of the largest off-diagonal entry of correlation matrices. *Trans. Amer. Math. Soc.* **359**, 5345–5363.